ISIS: A two-scale framework to sparse inference

Jianqing Fan
Princeton University

http://www.princeton.edu/~jqfan

March 26, 2010
1 Introduction
2 Challenge of Dimensionality
3 The ISIS Method
4 Sparse Survival analysis
5 Numerical Examples
High-dim variable selection characterizes many contemporary statistical problems.

- Biinformatic: disease classification / predicting clinical outcomes using microarray, proteomics, fMRI data;

- Association between phenotypes and SNPs or eQTL

- Document or text classification: E-mail spam.
- HPIs / drug sales collected in many regions
- Local correlation makes dimensionality growths quickly.

1000 neighborhoods requires 1 m parameters.
Managing 2K stocks involves 2m elements in covariance.
Example: Spatial temporal data

- Meteorology & Earth Sciences & Ecology
  - Temperatures and other attributes (precipitation, population size) are collected over time and over many regions.

- Large panel data over a short time horizon poses more challenges.
Growth of Dimensionality

- Dimen. grows rapidly w/ interactions: 5000 → 12.5m.

**Synergy of Two Genes**: colon cancer in Hanczar et al (2007).

e.g., \( Y = \mathbb{I}(X_1 + X_2 > 3) \) and \( Y \perp X_1 \).
Essential Assumption: Sparsity

**Dimensionality**: \( \log p = O(n^a) \) (NP-dimensionality)

**Intrinsic dimensionality**: \( s \ll n \). (Sparsity)
Impact of High-dimensionality
Impact of Dimensionality

**Regression:**

- Not directly implementable if $p > n$.
- Prediction error is $(1 + \frac{p}{n})\sigma^2$, if $p \leq n$.

**Classification:** No implementation problems, but error rates

—depend on $C_p^2 / \sqrt{p}$ (Fan & Fan 08), $C_p$ is distance.

—perfectly classifiable if $C_p^2 / \sqrt{p} \to \infty$ (Hall, Pittelkow & Ghosh, 08).
Impact of Dimensionality

Regression:
- **Not** directly implementable if $p > n$.
- Prediction error is $(1 + \frac{p}{n})\sigma^2$, if $p \leq n$.

Classification: No implementation problems, but error rates
- depend on $\frac{C_p^2}{\sqrt{p}}$ (Fan & Fan 08), $C_p$ is distance.
- perfectly classifiable if $\frac{C_p^2}{\sqrt{p}} \to \infty$ (Hall, Pittelkow & Ghosh, 08).
**dimensionality**: $p = 4500$, $n = 200$

**Signals**: $\mu_1 = 0.98\delta_0 + 0.02DE$, $\mu_2 = 0$
Impact of Dimensionality on classification

Classification power depends on \( \sum_{i=1}^{m} \alpha_i^2 / \sqrt{m} \).
An experiment: Generate $n = 50 \ Z_1, \cdots, Z_p \sim \text{i.i.d.} \ N(0, 1)$; compute $r = \max_{j \geq 2} \text{corr}(Z_1, Z_j)$.

\[ \begin{align*}
\text{compute maximum multiple correlation: } R &= \max_{|S|=5} \text{corr}(Z_1, Z_S).
\end{align*} \]
**An experiment**: Generate \( n = 50 \) \( Z_1, \cdots, Z_p \sim \text{i.i.d. } N(0, 1) \); compute 
\[
    r = \max_{j \geq 2} \text{corr}(Z_1, Z_j).
\]

\[
\begin{array}{c}
\text{Density} \\
\text{Maximum absolute sample correlation}
\end{array}
\]

\[
\begin{array}{c}
\text{Density} \\
\text{Maximum absolute multiple correlation}
\end{array}
\]

- compute maximum multiple correlation: 
\[
    R = \max_{|S| = 5} \text{corr}(Z_1, Z_S).
\]
**Scientific implication**: If $Z_1$ is responsible for breast cancer, but we can also discover other 5 genes, **indep of outcome**!

**False statistical inferences**: If $Y = Z_1$ and fit

$$Y = Z_{\hat{S}}^T \beta + \epsilon,$$

the residual variance

$$\hat{\sigma}^2 = \frac{RSS}{n - |\hat{S}|} \approx (1 - 0.9^2) \times \frac{49}{45} \approx 0.2,$$

★ $\hat{\sigma}$ is deflation by a factor of more than 1/2
★ more variables to be called “statistically significant”.
**Scientific implication**: If $Z_1$ is responsible for breast cancer, but we can also discover other 5 genes, **indep of outcome**!

**False statistical inferences**: If $Y = Z_1$ and fit

$$Y = Z_T^S \beta + \varepsilon,$$

the residual variance

$$\hat{\sigma}^2 = \frac{RSS}{n - |\hat{S}|} \approx (1 - 0.9^2) \times \frac{49}{45} \approx 0.2,$$

★ $\hat{\sigma}$ is deflation by a factor of more than 1/2

★ more variables to be called “statistically significant”.
**False Outcomes**

**Scientific implication:** If $Z_1$ is responsible for breast cancer, but we can also discover other 5 genes, **indep of outcome**!

**False statistical inferences:** If $Y = Z_1$ and fit

$$Y = Z^T\hat{S} \beta + \epsilon,$$

the residual variance

$$\hat{\sigma}^2 = \frac{RSS}{n - |\hat{S}|} \approx (1 - 0.9^2) \times \frac{49}{45} \approx 0.2,$$

★ $\hat{\sigma}$ is deflation by a factor of more than $1/2$

★ more variables to be called “statistically significant”.

Jianqing Fan (Princeton University)
Curse of Ultrahigh Dimensionality

- Computational cost
- Stability
- Estimation accuracy
- Noise Accumulation

Key Idea: Large-scale screening + moderate-scale searching.
The ISIS Method

a two-scale framework
**Indep learning**: Feature ranking by **Marginal** correlation (*Fan & Lv, 08*) or generalized correlation (*Hall & Miller, 09*);

**Classification**: Feature ranking by two-sample t-tests or other tests (*Tibshirani, et al, 03; Fan and Fan, 2008,*).
Extensions & Questions

Other methods: ★ Marginal LR (Fan, Samworth & Wu, 09);
★ MMLE (Fan and Song, 09); ★ MPLE (Zhao & Li, 09);
★ Nonparametric learning (Fan, Feng, Song, 09)
★ Data-tilting; (Hall, Titterington & Xue, 09).

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?
Extensions & Questions

**Other methods:** ★ Marginal LR \((Fan, Samworth & Wu, 09)\);
★ MMLE \((Fan and Song, 09)\); ★ MPLE \((Zhao & Li, 09)\);
★ Nonparametric learning \((Fan, Feng, Song, 09)\)
★ Data-tilting; \((Hall, Titterington & Xue, 09)\).

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?
Extensions & Questions

**Other methods:** ★ Marginal LR *(Fan, Samworth & Wu, 09)*;
★ MMLE *(Fan and Song, 09)*; ★ MPLE *(Zhao & Li, 09)*;
★ Nonparametric learning *(Fan, Feng, Song, 09)*
★ Data-tilting; *(Hall, Titterington & Xue, 09)*.

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?
Other methods: ★ Marginal LR (Fan, Samworth & Wu, 09);
★ MMLE (Fan and Song, 09); ★ MPLE (Zhao & Li, 09);
★ Nonparametric learning (Fan, Feng, Song, 09)
★ Data-tilting; (Hall, Titterington & Xue, 09).

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?
False Negative: What if $X_4$ marginally uncorrelated with $Y$, but jointly correlated with $Y$?

$$Y = X_1 + X_2 + X_3 + \beta_4 X_4 + \epsilon \quad \text{s.t.} \quad \text{cov}(Y, X_4) = 0.$$

False Positive: What if $X_2, \ldots, X_{99}$ highly correlated with an important $X_1$, but weakly correlated with $Y$ conditionally?

$$Y = X_1 + 0.2 X_{100} + \epsilon$$
Potential Drawbacks

♦ False Negative: What if $X_4$ marginally uncorrelated with $Y$, but jointly correlated with $Y$?

$$Y = X_1 + X_2 + X_3 + \beta_4 X_4 + \epsilon \quad \text{s.t.} \quad \text{cov}(Y, X_4) = 0.$$ 

♦ False Positive: What if $X_2, \cdots, X_{99}$ highly correlated with an important $X_1$, but weakly correlated with $Y$ conditionally?

$$Y = X_1 + 0.2 X_{100} + \epsilon$$
Oxygen Atom: Penalized likelihood estimation

\[ Q(\beta) = n^{-1} \sum_{i=1}^{n} L(Y_i, x_i^T \beta) + \sum_{j=1}^{d} p_\lambda(\|\beta_j\|) \]

Simultaneously estimate coefs and choose variables.

- How high dimensionality can such methods handle?
- What is the role of penalty functions?
- Does it possess an oracle property? How to compute?
Oxygen Atom: Penalized likelihood estimation

\[ Q(\beta) = n^{-1} \sum_{i=1}^{n} L(Y_i, x_i^T \beta) + \sum_{j=1}^{d} p_{\lambda}(|\beta_j|) \]

Simultaneously estimate coefs and choose variables.

- How high dimensionality can such methods handle?
- What is the role of penalty functions?
- Does it possess an oracle property? How to compute?
Penalized likelihood estimation

**Penalty**: Popular choice $L_1$. Preferred choice: SCAD or MCP.

Better bias property and model selection consistency.

$$Q^{\text{app}}(\beta) = n^{-1} \sum_{i=1}^{n} L(Y_i, x_{i,T} \beta) + \sum_{j=1}^{d} w_j |\beta_j|, \quad w_j = p'_{\lambda}(|\beta_j^{(k)}|)$$

**Oracle property**: can handle NP-dimensionality.

**Computation**: LQA (Fan & Li, 01); LLA (Zou & Li, 08); PLUS (Zhang, 09); Coordinate optimization (Fu & Jiang, 99).
Penalized likelihood estimation

**Penalty**: Popular choice $L_1$. Preferred choice: SCAD or MCP.

Better bias property and model selection consistency.

$$Q^{\text{app}}(\beta) = n^{-1} \sum_{i=1}^{n} L(Y_i, x^T_{i,d}\beta) + \sum_{j=1}^{d} w_j |\beta_j|,$$

$$w_j = p'_\lambda(||\beta_j^{(k)}||)$$

**Oracle property**: can handle NP-dimensionality.

**Computation**: LQA \((Fan & Li, 01)\); LLA \((Zou & Li, 08)\);

PLUS \((Zhang, 09)\); Coordinate optimization \((Fu & Jiang, 99)\).
Iterative application of large-scale screening and moderate-scale selection.

ISIS ((Fan & Lv, 08; Fan, Samworth & Wu, 09)), available in R.
Iterative conditional feature selection

1. **(Large-scale screening):** Apply SIS to pick a set $\mathcal{A}_1$;
   **(Moderate-scale selection):** Employ a penalized likelihood to select a subset $\mathcal{M}_1$ of these indices.

2. **(Large-scale screening):** Rank features according to the additional (conditional) contribution:

   $$
   L_j^{(2)} = \min_{\beta_0, \beta_{\mathcal{M}_1}, \beta_j} n^{-1} \sum_{i=1}^{n} L(Y_i, \beta_0 + x_{i, \mathcal{M}_1}^T \beta_{\mathcal{M}_1} + x_{ij} \beta_j),
   $$

   resulting in $\mathcal{A}_2$.

   — Improvement of residual approach (FL 08): $\beta_{\mathcal{M}_1} = \hat{\beta}_{\mathcal{M}_1}$.
Iterative conditional feature selection

1. **(Large-scale screening)**: Apply SIS to pick a set $\mathcal{A}_1$;
2. **(Moderate-scale selection)**: Employ a penalized likelihood to select a subset $\mathcal{M}_1$ of these indices.

(Large-scale screening): Rank features according to the additional (conditional) contribution:

$$L_j^{(2)} = \min_{\beta_0, \beta_{\mathcal{M}_1}, \beta_j} n^{-1} \sum_{i=1}^{n} L( Y_i, \beta_0 + x_{i,\mathcal{M}_1}^T \beta_{\mathcal{M}_1} + X_{ij} \beta_j ),$$

resulting in $\mathcal{A}_2$.

—Improvement of residual approach (FL 08): $\beta_{\mathcal{M}_1} = \hat{\beta}_{\mathcal{M}_1}$. 

Jianqing Fan  
*(Princeton University)*  
Sparse inference
Illustration of ISIS

All candidates

Conditional Screening

Moderate-scale selection

All candidates

Jianqing Fan (Princeton University)
(Moderate-scale selection): Minimize wrt $\beta_{M_1}, \beta_{A_2}$

$$
\sum_{i=1}^{n} L(Y_i, \beta_0 + x_{i,M_1}^T \beta_{M_1} + x_{i,A_2}^T \beta_{A_2}) + \sum_{j \in M_1 \cup A_2} p_\lambda(||\beta_j||),
$$

resulting in $M_2$

—Allow deletion, improvement over ISIS (Fan and Lv, 08).

Repeat Steps 1–3 until $|M_\ell| = d$ (prescribed) or $M_\ell = M_{\ell-1}$.
Iterative feature selection (II)

(Moderate-scale selection): Minimize wrt $\beta_{M_1}, \beta_{A_2}$

$$
\sum_{i=1}^{n} L(Y_i, \beta_0 + x_{i,M_1}^T \beta_{M_1} + x_{i,A_2}^T \beta_{A_2}) + \sum_{j \in M_1 \cup A_2} p_\lambda(||\beta_j||),
$$
resulting in $M_2$

—Allow deletion, improvement over ISIS (Fan and Lv, 08).

Repeat Steps 1–3 until $|M_\ell| = d$ (prescribed) or $M_\ell = M_{\ell-1}$.
The idea of ISIS is widely applicable. It can be applied to:

- **Classification** (Fan, Samworth, & Wu, 09).
- **Survival analysis** (Zhao & Li, 09; Fan, Feng, & Wu, 09).
- Nonparametric learning (Fan, Feng, & Song, 09).
- Robust and quantile regression (Bradic, Fan, & Wang, 09).
The idea of ISIS is widely applicable. It can be applied to:

- **Classification** (*Fan, Samworth, & Wu, 09*).

- **Survival analysis** (*Zhao & Li, 09; Fan, Feng, & Wu, 09*).

- Nonparametric learning (*Fan, Feng, & Song, 09*).

- Robust and quantile regression (*Bradic, Fan, & Wang, 09*).
The idea of ISIS is widely applicable. It can be applied to:

- Classification \((Fan, Samworth, & Wu, 09)\).

- **Survival analysis** \((Zhao & Li, 09; Fan, Feng, & Wu, 09)\).

- Nonparametric learning \((Fan, Feng, & Song, 09)\).

- Robust and quantile regression \((Bradic, Fan, & Wang, 09)\).
The idea of ISIS is widely applicable. It can be applied to

- **Classification** (*Fan, Samworth, & Wu, 09*).

- **Survival analysis** (*Zhao & Li, 09; Fan, Feng, & Wu, 09*).

- Nonparametric learning (*Fan, Feng, & Song, 09*).

- Robust and quantile regression (*Bradic, Fan, & Wang, 09*)
The idea of ISIS is widely applicable. It can be applied to

- **Classification** \((Fan, Samworth, & Wu, 09)\).

- **Survival analysis** \((Zhao & Li, 09; Fan, Feng, & Wu, 09)\).

- Nonparametric learning \((Fan, Feng, & Song, 09)\).

- Robust and quantile regression \((Bradic, Fan, & Wang, 09)\).
Sparse Survival Analysis
Hazard regression model

**Notation:**

- Observation time: $Y = \min(T, C)$.
- Censoring indicator $\delta = I(T \leq C)$.
- Covariates: $\mathbf{X} = (X_1, X_2, \cdots, X_p)^T$.

Cox’s proportional hazards model (Cox, 1972 & 1975):

$$h(t|\mathbf{x}) = h_0(t)\psi(\mathbf{x}) = h_0(t)\exp(\mathbf{x}^T\beta).$$

$\blacksquare$ $\beta$ sparse.
Hazard regression model

**Notation:**

- Observation time: $Y = \min(T, C)$
- Censoring indicator $\delta = I(T \leq C)$.
- Covariates: $X = (X_1, X_2, \cdots, X_p)^T$.

**Cox’s proportional hazards model** (Cox, 1972 & 1975):

$$h(t|x) = h_0(t)\psi(x) = h_0(t)\exp(x^T\beta).$$

$\blacksquare$ $\beta$ sparse.
**Marginal utility**: Screening according to $|\hat{\beta}_j^M|$ or

\[
u_j = \max_{\beta_j} \left( \sum_{i=1}^{n} \delta_i x_{ij} \beta_j - \sum_{i=1}^{n} \delta_i \log \left\{ \sum_{m \in \mathcal{R}(y_i)} \exp(x_{mj} \beta_j) \right\} \right).
\]

**Thresholding parameter**: Zhao and Li (09) proposed using upper $\alpha$ (Control of FDR) quantile of marginal utilities for decoupled response and covariate (PSIS).

**Theoretical properties**: sure independence screening property with false positive rate controlled by $\alpha$. (Zhao and Li, 09)
**Marginal partial likelihood**

**Marginal utility**: Screening according to $|\hat{\beta}_j^M|$ or

$$u_j = \max_{\hat{\beta}_j} \left( \sum_{i=1}^{n} \delta_i x_{ij}\hat{\beta}_j - \sum_{i=1}^{n} \delta_i \log \left\{ \sum_{m \in R(y_i)} \exp(x_{mj}\beta_j) \right\} \right).$$

**Thresholding parameter**: Zhao and Li (09) proposed using upper $\alpha$ (Control of FDR) quantile of marginal utilities for decoupled response and covariate (PSIS).

**Theoretical properties**: sure independence screening property with false positive rate controlled by $\alpha$. (Zhao and Li, 09)
Marginal partial likelihood

**Marginal utility**: Screening according to $|\hat{\beta}_j^M|$ or

$$u_j = \max_{\beta_j} \left( \sum_{i=1}^{n} \delta_i x_{ij} \beta_j - \sum_{i=1}^{n} \delta_i \log \left\{ \sum_{m \in \mathcal{R}(y_i)} \exp(x_{mj} \beta_j) \right\} \right).$$

**Thresholding parameter**: Zhao and Li (09) proposed using upper $\alpha$ (Control of FDR) quantile of marginal utilities for decoupled response and covariate (PSIS).

**Theoretical properties**: sure independence screening property with false positive rate controlled by $\alpha$. (Zhao and Li, 09)
Among active variables $I$ that pass the screening process, select them by **Penalized partial likelihood**:

$$
\min_{\beta_I} \left( -\sum_{i=1}^{n} \delta_i x_{I,i}^T \beta_I + \sum_{i=1}^{n} \delta_i \log\left\{ \sum_{m \in \mathcal{R}(y_i)} \exp(x_{m,I}^T \beta_I) \right\} + \sum_{j \in I} \rho_\lambda(\beta_j) \right),
$$

Reduce the active set further to $\hat{\mathcal{M}}_1$.

**Conditional screening**: All variables are candidates again. Rank them by conditional contribution given variables in $\hat{\mathcal{M}}_1$. 
Among active variables $I$ that pass the screening process, select them by **Penalized partial likelihood**:

$$
\min_{\beta_I} \left( -\sum_{i=1}^{n} \delta_i x_{I,i}^T \beta_I + \sum_{i=1}^{n} \delta_i \log \left\{ \sum_{m \in R(y_i)} \exp(x_{m,I}^T \beta_I) \right\} + \sum_{j \in I} p_\lambda(\beta_j) \right),
$$

Reduce the active set further to $\hat{M}_1$.

**Conditional screening**: All variables are candidates again. Rank them by conditional contribution given variables in $\hat{M}_1$. 
Simulation Studies

Fan, Feng and Wu (2010)
Simulation settings

- \( n = 300 \) and \( p = 400 \), baseline hazard \( h_0(t) = 0.1 \).

- Designs: \( X_1, \cdots, X_p \) are normal, equally correlated.
  - ★ Case 1: \( \rho = 0 \).
  - ★ Case 2: \( \rho = 0.5 \).
  - ★ Case 3: \( \rho = 1/\sqrt{2} \).
  - ★ Case 4: case 3 + an indep variable.

- Easy to difficulty: case 1 to case 4.

- **Hidden signature variable**: Large impact but indep. of outcome.
Simulation settings

- \( n = 300 \) and \( p = 400 \), baseline hazard \( h_0(t) = 0.1 \).

- Designs: \( X_1, \ldots, X_p \) are normal, equally correlated.
  - ★ Case 1: \( \rho = 0 \).
  - ★ Case 2: \( \rho = 0.5 \).
  - ★ Case 3: \( \rho = 1/\sqrt{2} \).
  - ★ Case 4: case 3 + an indep variable.

- Easy to difficulty: case 1 to case 4.

- **Hidden signature variable**: Large impact but indep. of outcome.
Simulation settings

- $n = 300$ and $p = 400$, baseline hazard $h_0(t) = 0.1$.

- Designs: $X_1, \cdots, X_p$ are normal, equally correlated.
  - ★ Case 1: $\rho = 0$.
  - ★ Case 2: $\rho = 0.5$.
  - ★ Case 3: $\rho = 1/\sqrt{2}$.
  - ★ Case 4: case 3 + an indep variable.

- Easy to difficulty: case 1 to case 4.

- **Hidden signature variable**: Large impact but indep. of outcome.
Design of Simulation

Case 1 & 2: \( \beta_1 = -1.63, \beta_2 = 1.34, \beta_3 = -1.65, \beta_4 = 1.65, \beta_5 = -1.42, \beta_6 = 1.70. \)

Case 3: \( \beta_1 = \beta_2 = \beta_3 = 4 \) and \( \beta_4 = -6\sqrt{2}. \)

**Hidden signature variable**: \( X_4, \text{ indep } X^{T}\beta^* \text{ and } T. \)

Case 4: \( \beta_1 = \beta_2 = \beta_3 = 4, \beta_4 = -6\sqrt{2} \) and \( \beta_5 = 4/3. \) **Weak variable**: \( X_5, \text{ marginally ranked below others.} \)

★ **Challenging**: Case 4 contains a hidden signature variable (false negative) and a weak variable (false positive).

★ Censoring rate: about 30%
Case 1 & 2: $\beta_1 = -1.63, \beta_2 = 1.34, \beta_3 = -1.65, \beta_4 = 1.65, \beta_5 = -1.42, \beta_6 = 1.70$.

Case 3: $\beta_1 = \beta_2 = \beta_3 = 4$ and $\beta_4 = -6\sqrt{2}$.

**Hidden signature variable:** $X_4$, indep $X^T\beta^*$ and $T$.

Case 4: $\beta_1 = \beta_2 = \beta_3 = 4$, $\beta_4 = -6\sqrt{2}$ and $\beta_5 = 4/3$.

**Weak variable:** $X_5$, marginally ranked below others.

★ **Challenging:** Case 4 contains a hidden signature variable (false negative) and a weak variable (false positive).

★ Censoring rate: about 30%
Design of Simulation

Case 1 & 2: $\beta_1 = -1.63, \beta_2 = 1.34, \beta_3 = -1.65, \beta_4 = 1.65, \beta_5 = -1.42, \beta_6 = 1.70$.

Case 3: $\beta_1 = \beta_2 = \beta_3 = 4$ and $\beta_4 = -6\sqrt{2}$.

**Hidden signature variable**: $X_4$, indep $X^T\beta^*$ and $T$.

Case 4: $\beta_1 = \beta_2 = \beta_3 = 4, \beta_4 = -6\sqrt{2}$ and $\beta_5 = 4/3$. **Weak variable**: $X_5$, marginally ranked below others.

★ **Challenging**: Case 4 contains a hidden signature variable (false negative) and a weak variable (false positive).

★ Censoring rate: about 30%
## Results for Cases 1 and 2 (100 simulations)

### Case 1: independent covariates

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>( | \beta - \hat{\beta} |_1 )</td>
<td>0.79</td>
<td>0.57</td>
<td>4.23</td>
</tr>
<tr>
<td>( | \beta - \hat{\beta} |_2 )</td>
<td>0.13</td>
<td>0.09</td>
<td>0.98</td>
</tr>
<tr>
<td>Proportion</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MMS</td>
<td>7</td>
<td>6</td>
<td>68.5</td>
</tr>
</tbody>
</table>

### Case 2: Equi-correlated covariates with \( \rho = 0.5 \)

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>( | \beta - \hat{\beta} |_1 )</td>
<td>2.2</td>
<td>0.64</td>
<td>4.38</td>
</tr>
<tr>
<td>( | \beta - \hat{\beta} |_2 )</td>
<td>1.74</td>
<td>0.11</td>
<td>0.98</td>
</tr>
<tr>
<td>Proportion</td>
<td>0.71</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MMS</td>
<td>7</td>
<td>6</td>
<td>57</td>
</tr>
</tbody>
</table>
### Results for Cases 1 and 2 (100 simulations)

#### Case 1: independent covariates

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\beta - \hat{\beta}|_1$</td>
<td>0.79</td>
<td>0.57</td>
<td>4.23</td>
</tr>
<tr>
<td>$|\beta - \hat{\beta}|_2$</td>
<td>0.13</td>
<td>0.09</td>
<td>0.98</td>
</tr>
<tr>
<td>Proportion</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MMS</td>
<td>7</td>
<td>6</td>
<td>68.5</td>
</tr>
</tbody>
</table>

#### Case 2: Equi-correlated covariates with $\rho = 0.5$

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\beta - \hat{\beta}|_1$</td>
<td>2.2</td>
<td>0.64</td>
<td>4.38</td>
</tr>
<tr>
<td>$|\beta - \hat{\beta}|_2$</td>
<td>1.74</td>
<td>0.11</td>
<td>0.98</td>
</tr>
<tr>
<td>Proportion</td>
<td>0.71</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MMS</td>
<td>7</td>
<td>6</td>
<td>57</td>
</tr>
</tbody>
</table>
## Results for Cases 3 and 4

### Case 3: A hidden signature variable

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\beta - \hat{\beta}|_1$</td>
<td>20.1</td>
<td>1.03</td>
<td>20.53</td>
</tr>
<tr>
<td>$|\beta - \hat{\beta}|_2^2$</td>
<td>94.72</td>
<td>0.49</td>
<td>76.31</td>
</tr>
<tr>
<td>Proportion</td>
<td>0</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>MMS</td>
<td>13</td>
<td>4</td>
<td>118.5</td>
</tr>
</tbody>
</table>

### Case 4: Two very hard variables to be selected.

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\beta - \hat{\beta}|_1$</td>
<td>20.87</td>
<td>1.15</td>
<td>21.04</td>
</tr>
<tr>
<td>$|\beta - \hat{\beta}|_2^2$</td>
<td>96.46</td>
<td>0.51</td>
<td>77.03</td>
</tr>
<tr>
<td>Proportion</td>
<td>0</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>MMS</td>
<td>13</td>
<td>5</td>
<td>118</td>
</tr>
</tbody>
</table>
## Results for Cases 3 and 4

### Case 3: A hidden signature variable

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\beta - \hat{\beta}|_1$</td>
<td>20.1</td>
<td>1.03</td>
<td>20.53</td>
</tr>
<tr>
<td>$|\beta - \hat{\beta}|_2$</td>
<td>94.72</td>
<td>0.49</td>
<td>76.31</td>
</tr>
<tr>
<td>Proportion</td>
<td>0</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>MMS</td>
<td>13</td>
<td>4</td>
<td>118.5</td>
</tr>
</tbody>
</table>

### Case 4: Two very hard variables to be selected.

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
<th>LASSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\beta - \hat{\beta}|_1$</td>
<td>20.87</td>
<td>1.15</td>
<td>21.04</td>
</tr>
<tr>
<td>$|\beta - \hat{\beta}|_2$</td>
<td>96.46</td>
<td>0.51</td>
<td>77.03</td>
</tr>
<tr>
<td>Proportion</td>
<td>0</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>MMS</td>
<td>13</td>
<td>5</td>
<td>118</td>
</tr>
</tbody>
</table>
## Computation costs

Average running time (in seconds) for Van-ISIS and LASSO.

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van-ISIS</td>
<td>379.29</td>
<td>213.44</td>
<td>402.94</td>
<td>231.68</td>
</tr>
<tr>
<td>LASSO</td>
<td>37730.82</td>
<td>26348.12</td>
<td>46847.00</td>
<td>28157.71</td>
</tr>
</tbody>
</table>
## Cases 2 and 4 with \( p = 1000 \)

### Case 2 with \( p = 1000 \) and \( n = 400 \)

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( | \beta - \hat{\beta} |_1 )</td>
<td>1.53</td>
<td>0.52</td>
</tr>
<tr>
<td>( | \beta - \hat{\beta} |_2 )</td>
<td>0.9</td>
<td>0.07</td>
</tr>
<tr>
<td>Proportion</td>
<td>0.82</td>
<td>1</td>
</tr>
<tr>
<td>MMS</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

### Case 4 with \( p = 1000 \) and \( n = 400 \)

<table>
<thead>
<tr>
<th></th>
<th>SIS</th>
<th>ISIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( | \beta - \hat{\beta} |_1 )</td>
<td>20.88</td>
<td>0.99</td>
</tr>
<tr>
<td>( | \beta - \hat{\beta} |_2 )</td>
<td>93.53</td>
<td>0.39</td>
</tr>
<tr>
<td>Proportion</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MMS</td>
<td>16</td>
<td>5</td>
</tr>
</tbody>
</table>
Cases 2 and 4 with $p = 1000$

<table>
<thead>
<tr>
<th></th>
<th>Case 2 with $p = 1000$ and $n = 400$</th>
<th></th>
<th>Case 4 with $p = 1000$ and $n = 400$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SIS</td>
<td>ISIS</td>
<td>SIS</td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\beta - \hat{\beta}</td>
<td></td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\beta - \hat{\beta}</td>
<td></td>
</tr>
<tr>
<td>Proportion</td>
<td>0.82</td>
<td>1</td>
<td>Proportion</td>
</tr>
<tr>
<td>MMS</td>
<td>8</td>
<td>6</td>
<td>MMS</td>
</tr>
</tbody>
</table>

Jianqing Fan (Princeton University) Sparse inference
An Application

Fan, Feng and Wu (2010)
Neuroblastoma data set

- Purpose: study genes related to survival information for German Neuroblastoma Trials.
- \( n = 251, p = 10,707, \) Affymetrix arrays.
- The censoring rate is 205/246.
- Survival curve of 246 patients is given by
## Selected Genes and its impact

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>estimated coefficient</th>
<th>standard error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_23_P31816</td>
<td>0.864</td>
<td>0.203</td>
<td>2.1e-05</td>
</tr>
<tr>
<td>A_23_P31816</td>
<td>-0.940</td>
<td>0.314</td>
<td>2.8e-03</td>
</tr>
<tr>
<td>A_23_P31816</td>
<td>-0.815</td>
<td>1.704</td>
<td>6.3e-01</td>
</tr>
<tr>
<td>A_32_P424973</td>
<td>-1.957</td>
<td>0.396</td>
<td>7.8e-07</td>
</tr>
<tr>
<td>A_32_P159651</td>
<td>-1.295</td>
<td>0.185</td>
<td>2.6e-12</td>
</tr>
<tr>
<td>Hs61272.2</td>
<td>1.664</td>
<td>0.249</td>
<td>2.3e-11</td>
</tr>
<tr>
<td>Hs13208.1</td>
<td>-0.789</td>
<td>0.149</td>
<td>1.1e-07</td>
</tr>
<tr>
<td>Hs150167.1</td>
<td>1.708</td>
<td>1.687</td>
<td>3.1e-01</td>
</tr>
</tbody>
</table>

- Six genes are significant and not removable.
- Adding two random selected genes 20 times, **no pairs** have significant contributions.
Estimated baseline survival function

Survival Probability

Years
Acknowledgement

Thank You

In collaboration with

★ Yang Feng (Princeton University; FFW, 2009)

★ Jinchi Lv (University of Southern California; Fan & Lv; 2008)

★ Richard Samworth (Cambridge University; FSW, 2009).

★ Rui Song (Colorado State University, Fan & Song, 2009).

★ Yichao Wu (North Carolina State University, FSW, 2009).